MILITARY TECHNICAL COLLEGE

CAIRO - EGYPT

GENERALIZED FISHER'S INFORMATION AND POINT ESTIMATION THEORY

BY

Fouad M. Abbady

Military Technical College

Cairo - Egypt

ABSTRACT

The aim of this paper is to introduce a generalized form of Fisher's information based on the concept of divergence of a sample of size n for different values of the population parameter. By using this generalization, we obtain a new form;less restricted, of cramér-Rao inequality. Some illustrative examples are presented.

0. INTRODUCTION:

The aim of this paper is to introduce a generalization of the classical statistical theory of point estimations based on Fisher's concept of information in a sample of size n using cramér-Rao inequality . As it is well known, applications of the classical point estimation theory based on cramér-Rao inequality require that the sample joint density function must satisfies some regularity conditions which are some what numerous and consequentely applications is so limited. If these conditions are not guaranteed we may come to serious mistakes. Decreasing these restrictions on the statistical model is achieved by generalizing Fisher's information which is based on the concept of divergence of a sample of size n for different values of the parameter to be estimated, this generalization permits a straight forward generalization to the classical cramér - Rao inequality . The new form is without the previous rather awkward regularity conditions.

The paper consists of two sections. Throuh-out section one formulation of the statistical model and demonstration of the basic results are given. Section two is devoted to the applications of the derived criteria of estimation.

1. MODEL FORMULATION AND GENERALIZED FISHER'S INFORMATION

Through this section, we shall consider the probability space $(\Omega, \mathfrak{S}, P_\theta)$, Ω is the fundamental probability set(sample space), $\mathfrak{S}$ is the segma algebra of subsets of Ω and P_θ is a probability function defined on $(\Omega, \mathfrak{S})$ which depen d on the parameter $\theta \in \Theta$, Θ is an open subset of R(the set of real numbers). Let $X_1, X_2, \ldots, X_n$ be n independent identically distributed, continuous random variables, each one defined on the probability space$(\Omega, \mathfrak{S}, p_\theta, \theta \in \Theta)$ having joint distribution function $F_o(\mathbf{X}; \theta)$ and joint density function $f_o(\mathbf{X}; \theta)$, $\mathbf{X} = (x_1, \ldots, x_n)$ $f(x_i; \theta)$, $F(x_i, \theta)$ are the density and distribution functions of X_i, $i = 1,2,\ldots,n$.

Let us consider the following definition (see [3])

Definition 1.1: The quantity

$$D_n(\theta', \theta'') = E_{\theta'}\left(1 - \frac{f_o(\mathbf{X}, \theta'')}{f_o(\mathbf{X}, \theta')}\right)^2$$

6

is called the divergence of a sample of size n for different values θ' and θ'' of the unknown parameter θ. For simplicity, we shall call $D_n(\theta',\theta'')$ by divergence only and denote $D_1(\theta',\theta'')$ by $D(\theta',\theta'')$.
In the definition of $D_n(\theta',\theta'')$ we understand, $0(1-\frac{0}{a})^2=0$ for $a\geq 0$ and $0(1-\frac{a}{0})=\infty$ for $a>0$. It is easy to see that $0\leq D_n(\theta',\theta'')\leq\infty$.

Proposition 1.2:

The divergence $D(\theta',\theta'')$ as defined before is equal to,

$$D_n(\theta,\theta'')=\int_{R^n}\frac{f_o(\mathbf{x},\theta'')}{f_o(\mathbf{x},\theta')}\,d\mathbf{x}-1$$

Proof:

From the definition of divergence, we have

$$D_n(\theta',\theta'')=E_{\theta'}\left(1-\frac{f_o(\mathbf{x};\theta'')}{f_o(\mathbf{x};\theta')}\right)^2$$

$$=\int_{R^n}\left(1-\frac{f_o(\mathbf{x};\theta'')}{f_o(\mathbf{x};\theta')}\right)^2 f_o(\mathbf{x};\theta')\,d\mathbf{x}$$

$$=\int_{R^n}\left(f_o(\mathbf{x},\theta')-2f_o(\mathbf{x},\theta'')+\frac{f_o^2(\mathbf{x};\theta'')}{f_o(\mathbf{x};\theta')}\right)d\mathbf{x}$$

Consequently,

$$D_n(\theta',\theta'')=\int_{R^n}\frac{f_o^2(\mathbf{x},\theta'')}{f_o(\mathbf{x},\theta')}\,d\mathbf{x}-1$$

Lemma 1.3

For our statistical structure,

$$D_n(\theta',\theta'')+1=(D(\theta',\theta'')+1)^n$$

Proof:

From the precedent proposition

$$D_n(\theta',\theta'')=\int_{R^n}\frac{f_o^2(\mathbf{x},\theta'')}{f_o(\mathbf{x},\theta')}\,d\mathbf{x}-1$$

then,

$$D_n(\theta',\theta'') + 1 = \int_{R^n} \frac{f_o^2(\mathbf{x} ; \theta'')}{f_o(\mathbf{x} ; \theta')} d\mathbf{x}$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{\infty} \left(\frac{f(x_i, \theta'')}{f(x_i, \theta')} \right)^2 f(x_{i,\theta'})\, dx_i$$

$$\therefore D_n(\theta',\theta'') + 1 = \left(\int_{-\infty}^{\infty} \frac{f^2(x , \theta'')}{f(x , \theta')} dx \right)^n$$

But

$$D(\theta',\theta'') + 1 = \int_{-\infty}^{\infty} \frac{f^2(x , \theta'')}{f(x , \theta')} dx$$

Thus,

$$D_n(\theta',\theta'') + 1 = (D(\theta',\theta'') + 1)^n$$

Definition: 1.4

The quantity

$$\therefore \quad I_n(\theta) = \underset{\mathcal{E}\to 0}{\underline{\mathrm{Lim}}} \frac{1}{\mathcal{E}^2} D_n(\theta, \theta + \mathcal{E})$$

is called the generalized Fisher's information in a sample of size n. where Lim denotes the left hand limit. $I_1(\theta)$ will be denoted simply by $I(\theta)$.

Theorem: 1.5

$$I_n(\theta) = n\, I(\theta)$$

Proof:

from Lemma 1.3

$$D_n(\theta,\theta') + 1 = (D(\theta,\theta') + 1)^n$$

Using the Binomial theorem, we get,

$$D_n(\theta,\theta')+1 = 1 + n\, D(\theta,\theta') + \ldots + \binom{n}{k} D^k(\theta, \theta') + \ldots + D^n(\theta,\theta').$$

replacing θ' by $\theta + \mathcal{E}$ and dividing both sides by $\mathcal{E}^2$, we have,

$$\frac{1}{\mathcal{E}^2} D_n(\theta,\theta+\mathcal{E}) = n\frac{1}{\mathcal{E}^2} D(\theta,\theta+\mathcal{E}) + \ldots + \binom{n}{k}\frac{1}{\mathcal{E}^2} D^k(\theta,\theta+\mathcal{E}) + \ldots + \frac{1}{\mathcal{E}^2} D^n(\theta,\theta+\mathcal{E})$$

Taking the limit from left of both sides, we get,

$$I_n(\theta) = n \lim_{\mathcal{E}\to 0} \frac{1}{\mathcal{E}^2} D(\theta,\theta+\mathcal{E}) + \ldots + \binom{n}{k} \lim_{\mathcal{E}\to 0}\frac{1}{\mathcal{E}^2} D^k(\theta,\theta+\mathcal{E}) + \ldots + \lim_{\mathcal{E}\to 0} \frac{1}{\mathcal{E}^2} D^n(\theta,\theta+\mathcal{E})$$

But since,

$$\lim_{\mathcal{E}\to 0} \frac{1}{\mathcal{E}^2} D^k(\theta,\theta+\mathcal{E}) = 0 \quad \text{if } k > 1 \text{ and } 0 \le I(\theta) < \infty \text{, then}$$

$$I_n(\theta) = n\, I(\theta)$$

if $I(\theta) = \infty$, then $I_n(\theta) = \infty$.

__Theorem:__

The generalized information $I(\theta)$ as defined before is the same as the classical definition,

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{f'(x,\theta)}{f(x,\theta)}\right)^2 f(x,\theta)\,dx$$

provided the following conditions are satisfied

(1) $\frac{\partial}{\partial\theta} f(x;\theta)$ $(f'(x,\theta))$ exists for all $x \in R$

(2) $\int_{-\infty}^{\infty} \left(\frac{f'(x,\theta)}{f(x,\theta)}\right)^2 f(x,\theta) \quad \forall\ \theta$

(3) $\lim_{\mathcal{E}\to 0} \int_{-\infty}^{\infty} \left(\frac{f(x;\theta+\mathcal{E}) - f(x;\theta)}{\mathcal{E}\, f(x,\theta)} - \frac{f'(x;\theta)}{f(x;\theta)}\right)^2 f(x;\theta)\,dx = 0$

for any $\theta \in \Theta$.

Proof:

We shall start the proof by the following lemma.

Lemma. If we have $\varphi(x,\varepsilon)$, $\varphi(x)$ such that,

$$\int_{-\infty}^{\infty} \varphi^2(x) \; f(x,\theta) \; dx < \infty \text{ and}$$

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} (\varphi(x,\varepsilon) - \varphi(x))^2 \, f(x,\theta) \, dx = 0$$

then

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \varphi^2(x,\varepsilon) \, f(x;\theta) dx = \int_{-\infty}^{\infty} \varphi^2(x) \, f(x;\theta) \, dx$$

Proof:

Let us adopt the following notation,

$$\|\varphi(x;\varepsilon)\|^2 = \int_{-\infty}^{\infty} \varphi^2(x;\varepsilon) f(x;\theta) \, dx$$

$$\|\varphi(x)\|^2 = \int_{-\infty}^{\infty} \varphi^2(x) \, f(x;\theta) \, dx$$

we know that,

$$\|\varphi(x;\varepsilon)\| \leqq \|\varphi(x;\varepsilon) - \varphi(x)\| + \|\varphi(x)\| \quad \ldots \quad (1)$$

and

$$\big|\, \|\varphi(x;\varepsilon)\| - \|\varphi(x)\| \,\big| \leqq \|\varphi(x;\varepsilon) - \varphi(x)\| \quad \ldots \quad (2)$$

then,

$$\big|\, \|\varphi(x;\varepsilon)\|^2 - \|\varphi(x)\|^2 \big| = \big|\, \|\varphi(x;\varepsilon)\| - \|\varphi(x)\| \,\big| \cdot (\|\varphi(x;\varepsilon)\| + \|\varphi(x)\|)$$

using (1) and (2) we have,

$$\big|\, \|\varphi(x;\varepsilon)\|^2 - \|\varphi(x)\|^2 \big| \leqq \|\varphi(x,\varepsilon) - \varphi(x)\| \cdot (\|\varphi(x;\varepsilon) - \varphi(x)\| + 2\|\varphi(x)\|)$$

$$= \|\varphi(x,\varepsilon) - \varphi(x)\|^2 + 2 \, \|\varphi(x)\| \cdot \|\varphi(x;\varepsilon) - \varphi(x)\|$$

But since, $\|\varphi(x)\| < \infty$ and

$$\|\varphi(x;\varepsilon) - \varphi(x)\| \xrightarrow[\varepsilon \to 0]{} 0$$

so,

$$\left| \|\varphi(x;\varepsilon)\|^2 - \|\varphi(x)\|^2 \right| \xrightarrow[\varepsilon \to 0]{} 0$$

and therefore the lemma is proved.

If we denote , $\dfrac{f(x;\theta+\varepsilon) - f(x;\theta)}{\varepsilon \, f(x;\theta)}$ by $\varphi(x;\varepsilon)$

and $\dfrac{f'(x;\theta)}{f(x;\theta)}$ by $\varphi(x)$, then ,from condition (3) and using the above lemma, we have $\lim\limits_{\varepsilon\to 0} \left| \int_{-\infty}^{\infty} (\varphi^2(x;\varepsilon) - \varphi^2(x))\, f(x;\theta)\, dx \right| = 0$

i.e.,

$$\lim_{\varepsilon \to 0} \left| \|\varphi(x;\varepsilon)\|^2 - \|\varphi(x)\|^2 \right| = 0$$

therefore,

$$\lim_{\varepsilon\to 0} \int_{-\infty}^{\infty} \left(\frac{f(x;\theta+\varepsilon) - f(x;\theta)}{\varepsilon \, f(x;\theta)}\right)^2 f(x;\theta)\, dx = \int_{-\infty}^{\infty} \left(\frac{f'(x;\theta)}{f(x;\theta)}\right)^2 f(x;\theta)\,dx.$$

Hence,

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D(\theta;\theta+\varepsilon) = \int_{-\infty}^{\infty} \left(\frac{f'(x,\theta)}{f(x;\theta)}\right)^2 f(x;\theta)\,dx$$

which completes the proof.

Remark:

Due to the fact that $I_n(\theta) = n\, I(\theta)$, the above theorem is also true for $n > 1$.

Theorem:

For our statistical structure and for any unbiased estimate $T(x_1,\ldots,x_n)$ of the unknown parameter θ, we have

$$\operatorname{var}(T(x_1,\ldots,x_n)) \geqq \frac{1}{n\, I(\theta)} .$$

Proof:

The basic tool of the proof is schwarz inequaltiy,hence,

$$\left(\int_{R^n} (T(\mathbf{x}) - \theta)\left(\frac{f(\mathbf{x};\theta+\varepsilon) - f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)}\right) f(\mathbf{x};\theta)\, d\mathbf{x}\right)^2$$

$$\leqq \int_{R^n} (T(\mathbf{x})-\theta)^2 f(\mathbf{x};\theta)\, d\mathbf{x} \int_{R^n} \left(\frac{f(\mathbf{x};\theta+\varepsilon) - f(\mathbf{x};\theta)}{f(\mathbf{x},\theta)}\right)^2 f(\mathbf{x};\theta)\, d\mathbf{x}$$

$$\left(\int_{R^n}(T(\mathbf{X})-\theta)f(\mathbf{X};\theta+\varepsilon)\,d\mathbf{X}-\int_{R^n}(T(\mathbf{X})-\theta)f(\mathbf{X};\theta)\,d\mathbf{X}\right)^2 \leq \operatorname{var}(T(\mathbf{X}))\, D_n(\theta;\theta+\varepsilon)$$

then,

$$\operatorname{var}(T(\mathbf{X})) \geq \frac{1}{\frac{1}{\varepsilon^2} D_n(\theta;\theta+\varepsilon)}$$

taking $\lim\limits_{\varepsilon \to 0}$ of both sides , we have,

$$\operatorname{var}(T(\mathbf{X})) \geq \frac{1}{\lim\limits_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D_n(\theta;\theta+\varepsilon)} = \frac{1}{I_n(\theta)} = \frac{1}{n\, I(\theta)}$$

Therefore,

$$\operatorname{Var}(T(\mathbf{X})) \geq \frac{1}{n\, I(\theta)} .$$

2. APPLICATIONS

Through-out this section the derived result will be applied into three different statistical structures.

Example 1: The normal case.

$$f(x,\theta) = (2\pi)^{-\frac{1}{2}}\, e^{-\frac{(x-\theta)^2}{-2}}$$

$$\frac{f^2(x;\theta+\varepsilon)}{f(x;\theta)} = (2\pi)^{-\frac{1}{2}}\, e^{\varepsilon^2}\, e^{-\frac{1}{2}(x-(\theta+2\varepsilon))^2}$$

so,

$$D(\theta,\theta+\varepsilon)+1 = (2\pi)^{-\frac{1}{2}}\, e^{\varepsilon^2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-(\theta+2\varepsilon))^2}\, dx$$

Then,

$$D(\theta,\theta+\varepsilon)+1 = e^{\varepsilon^2}$$

consequently,

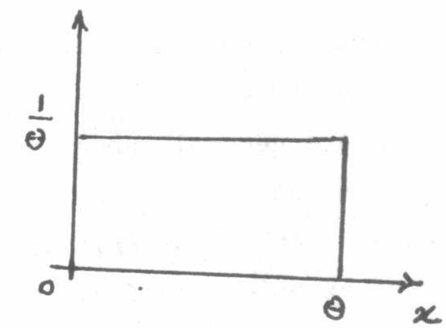
$$I(\theta) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2} D(\theta;\theta+\varepsilon) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\left(e^{\varepsilon^2}-1\right) = 1$$

therefore the minimum variance is $\frac{1}{n}$.

6

Example 2: **The uniform case**

$$f(x;\theta) = \frac{1}{\theta} \qquad 0 < x < \theta$$



$$\frac{f^2(x\,;\,\theta+\varepsilon)}{f(x\,;\,\theta)} = \frac{\theta}{(\theta+\varepsilon)^2}$$

Then,

$$D(\theta,\theta+\varepsilon) + 1 = \int_0^{\theta} \frac{\theta}{(\theta+\varepsilon)^2}\,dx$$

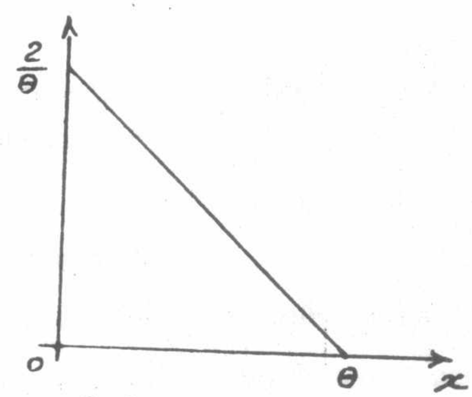
$$= \frac{\theta^2}{(\theta+\varepsilon)^2}$$

Thus

$$D(\theta;\theta+\varepsilon) = \left(\frac{\theta}{\theta+\varepsilon}\right)^2 - 1$$

$$I(\theta) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2}\, D(\theta;\theta+\varepsilon) = \infty$$

The minimum variance is therefore zero.



$0 < x < \theta$

Example 3: **The triangle case.**

$$f(x,\theta) = -\frac{2}{\theta^2}(x-\theta)$$

$$\frac{f^2(x;\theta+\varepsilon)}{f(x\,;\,\theta)} = -\frac{2\theta^2}{(\theta+\varepsilon)^4}\,\frac{(x-(\theta+\varepsilon))^2}{(x-\theta)}$$

$$= -\frac{2\theta^2}{(\theta+\varepsilon)^2}\left(x-(\theta+2\varepsilon) + \frac{\varepsilon^2}{x-\theta}\right).$$

Then

$$D(\theta,\theta+\varepsilon) + 1 = \frac{-2\theta^2}{(\theta+\varepsilon)^2}\int_0^{\theta}\left(x-(\theta+2\varepsilon) + \frac{\varepsilon^2}{x-\theta}\right)dx$$

Hence,

$$D(\theta,\theta+\varepsilon) = \infty \quad \text{and} \quad I(\theta) = \infty$$

therefore the minimum variance is zero.

Comment

It was shown through-out the paper that the classical form of Cramér-Rao inequality used in the books of mathematical statistics is a special case of more general concept.

It is easy to see that the classical form of Cramér- Rao inequality can be applied in the case of example 1 but it can not be applied for the cases of examples 2 and 3. unfortunately the minimum variance calculated using the generalized concept in the last two cases was zero. Investigation therefore should be continued to find models where the classical form can not be applied and the minimum variance is different from zero.

References

[1] J.Neveu (1965) : Mathematical Foundations of the calculus of Probability. Holden Day,san Francisco. U.S.

[2] C.R.Rao (1973) : Linear statistical inference and its Applications. second edition. John wiley & sons,New York .

[3] W. Rudin(1960) : Principles of Mathematical Analysis. Hill,Tokyo.

[4] H.G.Tucher(1965): An introduction To Probability and Mathematical statistics.Acadimic Press. New York.

[5] I.Vajda(1972) : On the f - divergence and singularity of probability measures. Periodics mathematica Hungarica Vol.2 pp.223 - 234.

[6] B.L. Vander Waerden(1969): Mathematical statistics, George Allen,London.